Catalogue no. 68-515-X
ISBN 978-0-660-02486-8

Integrated Business Statistics Program Overview

Release date: June 17, 2015

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at infostats@statcan.gc.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-877-287-4369

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "About us" > "The agency" > "Providing services to Canadians."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

. not available for any reference period
.. not available for a specific reference period
... not applicable
0 true zero or a value rounded to zero
0[s] value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
[p] preliminary
[r] revised
x suppressed to meet the confidentiality requirements of the *Statistics Act*
[E] use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2015

All rights reserved. Use of this publication is governed by the Statistics Canada Open Licence Agreement.

An HTML version **is also available.**

Cette publication est aussi disponible en français.

Table of Contents

1. Introduction and Background 4
2. IBSP Objectives and Guiding Principles 5
3. Cornerstones of an Integrated Infrastructure System 6
   a) Metadata driven model 6
   b) Building an integrated infrastructure 7
   c) Information management functionality 8
4. A Common Frame – The Business Register 9
   a) Source of data for allocation 9
   b) Source of information on commodities and activities 10
5. Developing a Harmonized Content Model 11
6. Managing Response Burden 12
   a) Tax replacement strategy 12
   b) Smart replacement strategy 12
   c) Active collection management 13
7. Incorporating a Coherence Analysis Framework 14
8. Large and Complex Enterprises – A Special Case 14
   a) EPM-LAOS programs 14
   b) Customized collection 15
9. Methodology Approach 15
   a) Two-phase sampling 15
   b) Calibration 16
10. Strategy for Analysis and Dissemination 17
   a) Analytic strategy 17
   b) Integration with the System of National Accounts (SNA) 18
11. Governance Model 19
   a) Committee structure 19
   b) Management information system 20
12. Conclusion 20
13. References 21

1. Introduction and Background

In 2010, Statistics Canada launched the Corporate Business Architecture (CBA) initiative. At the time, there were growing financial pressures in the organization which led to a thorough review of business methods, statistical processes, and systems infrastructure. The main objectives were to identify opportunities for efficiencies, determine methods for enhancing quality assurance, and finding ways to improve responsiveness in the delivery of statistical programs.

The work resulted in numerous recommendations including the development and mandatory use of shared and generic corporate services for collecting, processing, storing and disseminating statistical information. To achieve the goals identified, Statistics Canada initiated several projects including a major transformation project for its economic statistics surveys, the Integrated Business Statistics Program (IBSP).

The IBSP provides a standardized framework for economic surveys conducted at Statistics Canada. IBSP surveys use Statistics Canada's Business Register as a common frame. Questionnaires are based on harmonized concepts and content. And surveys share common sampling, collection and processing methodologies that are driven by metadata. In addition, common tools are in place to edit, correct, and analyse data.

Although the IBSP was an ambitious undertaking, the new program was not developed from scratch. In fact, it is a continuation of an effort to build a harmonized business surveying approach that began in the late 1990s with the Unified Enterprise Statistics (UES) program.[1] The UES program originally covered seven pilot surveys and gradually expanded to include sixty annual business surveys in the agriculture, manufacturing, trade and services sectors. But, over time, the UES systems infrastructure became antiquated. Substantial resources were required for systems maintenance and the model could not easily adapt to changing requirements. Thus, it was an opportune time to redesign the model and implement CBA principles.

Under IBSP, the business surveying infrastructure has been completely redeveloped and new innovative methodologies and processes have been introduced to improve upon the UES model. Many of these innovations resolve longstanding shortcomings of the UES, including implementing a system that has flexibility to adapt to new requirements. By 2016, the majority of Statistics Canada's economic surveys will be incorporated into the IBSP model.

This overview document describes various aspects of the IBSP and is intended to serve as an introduction to the model for current and future staff. It is one component of a comprehensive learning framework that includes detailed documentation describing methods and processes, user manuals that accompany tools, documents describing editing and analytic strategies, and classroom training sessions that afford staff the opportunity to work with IBSP tools and systems in a training environment before being asked to do so in production.

1. Refer to Brodeur et al. paper in references.

2. IBSP Objectives and Guiding Principles

In constructing the IBSP model, the team focused on six core objectives.

These were:

- improving data quality by applying standardised methods and processes, implementing harmonised content, and facilitating coherence analysis;
- reducing response burden;
- modernizing data processing infrastructure;
- integrating the majority of economic surveys into the new model;
- simplifying and standardizing processes to reduce learning curves and improve timeliness;
- reducing ongoing costs associated with operational aspects of surveys to realize efficiencies.

To attain these objectives, survey programs had to adapt to specific requirements of the IBSP model, but at the same time, the model had to be designed with flexibility to respond to unique program requirements. Achieving the right balance between developing a standardized, coherent model, while retaining flexibility for program specific requirements was the greatest challenge faced in implementing IBSP.

Common features of IBSP surveys that promote operational efficiency include:

- full use of the Business Register as the frame;
- use of electronic questionnaires as the principal mode of collection;
- active collection management based on quality indicators;
- implementation of a common editing strategy;
- applying standardized metadata for questionnaire development, sampling, edit and imputation, allocation and estimation processes;
- shared governance across operational and subject matter divisions including change management.

IBSP surveys use these common features, but there are many variants to accommodate survey specific requirements. For example, the model is designed with flexibility to process surveys with different frequencies, including monthlies, quarterlies and annuals. And with different coverage, such as economy wide surveys, industry based surveys and activity based surveys.

3. Cornerstones of an Integrated Infrastructure System

From its inception, the IBSP vision was to build a scalable and efficient business surveying infrastructure. In terms of scalability, the system was designed to incorporate a myriad of different surveys while minimizing processing constraints. For example, the system can process data simultaneously for multiple surveys, something that was not possible in the UES.

The system was also designed so that changes to requirements, including adding new surveys or changing specifications for existing surveys, could be implemented without putting inordinate strain on processing staff. Scalability and efficiency were achieved by designing a system that is driven by metadata. In the integrated metadata IBSP system, all processes can be run systematically without manual interventions to transform files as a survey progresses from one process to the next.

a) Metadata driven model

Statistics Canada has a long history of developing corporate metadata repositories for managing publications, services, and statistical holdings. However, relatively few survey programs have well developed metadata repositories for managing survey operations. The UES program did implement a metadata system that housed processing edits along with variable cell numbers and cell descriptions. For IBSP, this metadata framework has been expanded to cover all aspects of survey processing.[2] This approach increases efficiency, robustness, and responsiveness in delivering processing services for IBSP programs.

In the IBSP model, metadata are stored in easily modifiable tables that are used to drive systems programs. This is a departure from the UES model where metadata were often hard coded into programs. IBSP systems programs simply access information from metadata tables to direct their execution.

A key advantage of the IBSP metadata driven system is that changes required as program needs evolve can be accommodated by modifying metadata, rather than by rewriting system code. This provides more control for the processing team and more flexibility for users.

According to metadata management guidelines, metadata should be active, created for a purpose and used in downstream processes. While the 'no data without metadata' principle is often applied to final data output in the form of descriptive metadata, this is also true for processing, especially in the IBSP. When a variable is created, it is tagged with descriptive elements such as a name and an origin but metadata will also indicate how validation, editing and imputation must be done, and will track the variable's passage through the various processing steps.

Users will have a single point of entry into the IBSP, since metadata will be integrated into every processing step, and their management along with the processes they direct will naturally form components of the same seamless portal. This integration enables the system to instantly check that run conditions are met. For example, if a user chooses to execute a process, the interface can prompt the user to input the necessary metadata and ensure that other prerequisites are in place. Inputs can then be validated automatically, and as a result, the system would either stop to give a warning or allow the user to proceed with subsequent steps. The metadata interface approach does not require the user to deal with multiple applications and does not require knowledge about the order of the various steps necessary to run processing.

While the efficiency goal of a metadata driven system is to minimize rework and facilitate reuse, improving quality and coherence is an equally important outcome. The integration of metadata in processing operations facilitates automatic coherence checks. Data integrity rules are enforced through the system's database to ensure the quality of inputs. Metadata also generates invaluable management information to aid in monitoring progress, thereby improving the overall quality of survey processing.

2. Refer to Hostetter paper in references.

b) Building an integrated infrastructure

The IBSP infrastructure's core components are the suite of generic services required to perform sampling (G-SAM), edit and imputation (BANFF), estimation (G-EST), confidentiality (G-CONFID) and dissemination (G-EXPORT). In addition, a number of repositories for the different types of data, paradata and metadata that are used throughout the process act as Data Service Centres.

The main Data Service Centres are: the Business Register which serves as the frame, Tax warehouse which contains all tax files; the Integrated Metadata Base (IMDB) which contains metadata related to content; the IBSP data mart which contains all data processing files; and the Collection warehouse which contains raw data from respondents.

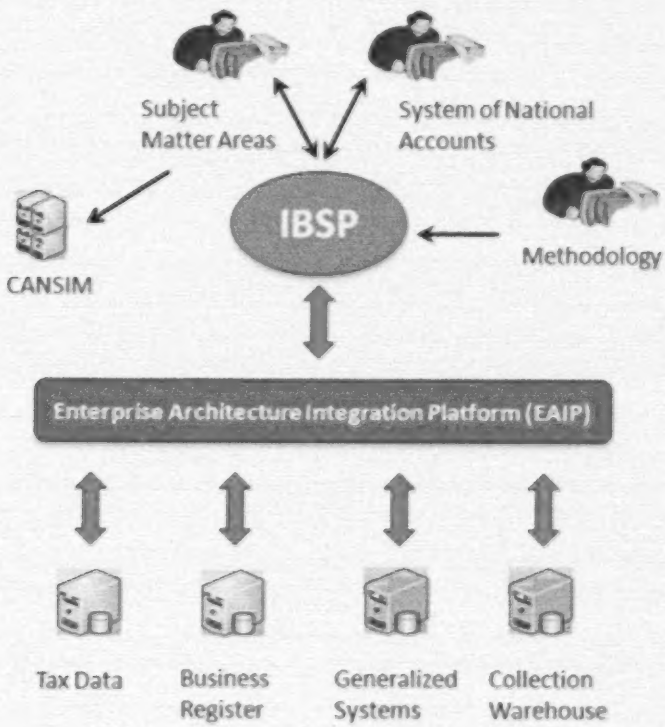
The integration of generic services is facilitated by using the Enterprise Architecture Integration Platform (EAIP) which allows seamless data transformations between core business services including:

- Business Register data services, Tax data services, IMDB metadata and, in the future, Collection data;
- Generic services for data processing (G-SAM, BANFF, G-EST).

One important benefit from adopting EAIP is that a modification to a data service by any given data service center will not force the IBSP to adapt its interfaces to accommodate the change. This promotes stability within the system and improves reliability.

The diagram (Figure 1) illustrates that data services are connected directly to EAIP. The technology then allows all connected parties to access requisite data files or processing services.

Figure 1
IBSP Infrastructure



The IBSP Technology Architecture (IBSP TA) supports the integrated infrastructure. It is based on the logical software and hardware capabilities required to support deployment of business, data, and application services. This includes IT infrastructure, middleware, networks, communications, processing, and standards.

The IBSP TA represents a formal description of the IBSP system, the structure of components, their inter-relationships, and the principles and guidelines governing their design and evolution over time. It also includes the major integration components with which the new system needs to interact and how the architecture will be implemented to accommodate these significant external components.

The IBSP TA delivers a metadata driven system heavily integrated with core Statistics Canada data services. Key attributes include:

- Scalability / Maintainability / Extensibility
  - layered architecture
  - service oriented architecture and EAIP
  - object oriented architecture
- Security
  - security groups based on roles
  - authentication through user account login
  - auditing mechanism to track user activities
  - no direct access to database tables
- Usability
  - iterative architecture
  - service oriented architecture
- Performance / Availability / Reliability
  - enterprise architecture, EAIP
  - corporate infrastructure services
  - corporate hosting services

c) Information management functionality

Applying sound management principles to statistical holdings is an integral component of Statistics Canada's work to provide Canadians with access to a trusted source of information. For IBSP, the integrated infrastructure is designed to facilitate management of data holdings based on both legislative requirements and Statistics Canada's information management directives.

One key aspect of information management is to maintain documentation that provides context for micro data and aggregate data files. This includes data dictionaries, definitions of concepts, and other supporting documents that are required by users who need to understand strengths, limitations and intended uses of the information. The level of documentation can vary, depending on the long term value assigned to information holdings. For IBSP, the bulk of documentation required to meet information management objectives will be created, maintained and retained in either the metadata framework or the IBSP data mart where files reside. Users will be able to access data files and associated metadata using standardized tools.

Another important aspect of information management is the application of retention rules to information holdings, and the deletion of information that has no business value. For IBSP, metadata identifying retention periods will be incorporated into data files to facilitate their management. Files no longer required will be overwritten or deleted via automated processes.

4. A Common Frame – The Business Register

The Business Register (BR) is the common frame for all surveys using the IBSP model. The BR is a data service centre updated through a number of sources including administrative data files, feedback received from conducting Statistics Canada business surveys, and profiling activities including direct contact with companies to obtain information about their operations and Internet research findings. Using the BR will ensure quality, while avoiding overlap between surveys and minimizing response burden to the greatest extent possible.

The BR identifies all businesses operating in Canada and foreign businesses that have links to Canadian companies. It includes information about how businesses are organized, the industries they operate in, their size in terms of revenues earned and number of employees, and their location.

The vast majority of businesses on the BR are referred to as simple businesses in that they have one operating entity and are classified to one industry and one geographic region. These simple businesses account for 99% of all businesses on the BR. Complex businesses, i.e., businesses that have more than one operating entity are small in number, 1% of the total, but they account for approximately 50% of revenues generated. The structure of complex businesses is mainly updated through profiling and survey feedback. Administrative data are also used but these data are often not available at the level required to update the full structure of complex business.

There are four statistical attributes that are derived for each business on the BR. The highest level is the statistical enterprise which often corresponds to a legal unit, but for some very complex businesses, it may consist of a family of legal units. The next level in the structure is the statistical company, which is the lowest level where investment can be measured. Operating profit can also be calculated at this level, and assets and liabilities can be measured.

Below the company is the statistical establishment, which in most cases is equivalent to a profit centre. At this level, the value of output, the cost of inputs, and labour can be measured. Finally, the statistical location is the lowest level entity. Employment and/or revenue data are available at this level.

IBSP surveys will target different levels in the business structure depending on their requirements. For example, for surveys measuring assets and liabilities, the statistical enterprise or company will be targeted. By contrast, many production surveys will target the statistical establishment or location since these levels provide the most detailed view of the industrial and geographic distribution of economic activity in Canada.

a) Source of data for allocation

To produce economic data at the most detailed level of a business's structure, which is a requirement for compiling national accounts, an allocation process is sometimes required. In general, the process takes administrative or survey data that is only available at the statistical company or enterprise level, and allocates the data to the business's various establishments and locations, so that provincial/territorial statistical distributions can be created. Over time, programs developed their own allocation methods to meet specific needs and allocation processes became diverse.

In moving towards standardised processes and methods under IBSP, allocation strategies were reviewed and a common method was developed and implemented. Benefits from adopting a single allocation methodology include improved coherence of survey and administrative data, and a more simplified integration of the allocation method into survey processing.

Because of its status as a centralized tool, its accessibility and its extensive use within the organization, it was decided that the BR would be the best place to store the allocation variables, which are know as allocation factors. The information used to derive the allocation factors comes in general from two main sources: profile variables on the BR; and administrative data from the Canada Revenue Agency T4 Supplemental file, which provides the distribution of salaries by province and territory. Together, these data are used to compile two sets of factors: revenue factors and employment factors. They show, respectively, the share of the enterprise's total revenue or employment belonging to each of its operating entities.

The allocation factors are first determined by using T4 data, provided the data satisfy the following three criteria:

- they are available for the business in question as referenced by the business number (BN);
- they cover all provinces and territories in which the legal entity has at least one operating entity;
- they are more current than the profile variables that reside on the BR.

Under these conditions, the enterprise allocators are derived such that the distribution of its employment and revenues is aligned with the provincial/territorial distribution shown by the T4 wages and salaries data.

If the T4 data do not satisfy the above-mentioned criteria or if all the operating entities are located in the same province or territory, the allocation factors are calculated directly from the profile variables available on the BR. Profiling is the process by which specialized analysts regularly survey enterprises to ensure that the BR has an up-to-date listing of the operating entities for a given enterprise. As part of this exercise, an estimate of the total employment and revenues is obtained for each entity. For the most complex businesses operating in Canada, profile data are updated frequently and, as a result, are most likely to be used to create the BR allocation factors.

For the majority of IBSP surveys, the BR revenue allocation factor is used to allocate financial, commodity and characteristics data in instances where collected data encompass national operations of complex businesses. The one exception is for the salaries and wages paid variable, where the BR employment allocation factor is applied.

As part of Statistics Canada's quality assurance process, the allocation factors are reviewed and validated by subject matter analysts. During this review, updates can occur. These updates are often based on enhanced information available to analysts such as information from data confrontation sources. Once validated by analysts, the allocation factors are applied and estimates are tabulated. For IBSP programs, estimates are generally compiled at a detailed industry level for each province/territory in Canada.

The extent of allocation varies across survey programs. It will depend on: the structure of a given industry, and specifically the number of complex businesses in the population; the breadth of data reported by respondents; and the collection strategy employed. For example, Statistics Canada may collect data from businesses for their national operations, their provincial operations, or all of their business locations. When data are collected for national operations, the impact of allocation will be greater. By contrast if data are separately reported for each province or territory where a business has operations, there is no requirement to apply allocation factors to further disaggregate the data.

The collection strategy selected by individual programs is based on output requirements, such as the level of precision required for the various estimates produced. Information about the collection strategy for any given industry is made available to users when data are publicly released.

b) Source of information on commodities and activities

To improve the quality of commodity estimates and to more efficiently select samples, survey programs identified a requirement to target businesses producing specific commodities as well as those engaged in specific activities. For example, in conducting a survey on capital expenditures, a more efficient sample can be selected if there is information available on the survey frame to flag which businesses have capital expenditures.

The BR now has capacity to retain commodity or activity information through the additional production and size measures field. Subject matter divisions are responsible for populating and maintaining the field. Once the information is available, it can be used during sample selection to target businesses that will significantly impact on the quality of commodity and/or activity estimates. For IBSP, several programs will utilise this functionality. In fact, the phase 1 sample of the IBSP, which is described later, will be used to collect commodity and activity information which will then be used to efficiently select the businesses that will receive a more detailed phase 2 survey questionnaire (see section 9).

5. Developing a Harmonized Content Model

Building an integrated infrastructure certainly helps in achieving operational efficiency, but the full benefits of an integrated approach can only be obtained by also implementing a harmonized conceptual framework. For IBSP, this begins with the application of standards. All IBSP surveys must apply statistical standards including:

- the North American Industrial Classification System (NAICS) to classify the target population by industry;
- the North American Product Classification System (NAPCS) to categorize and collect business input and output data;
- the Chart of Accounts (COA)[3] as the reference taxonomy for organizing business financial information. (e.g. revenue, expenses, assets and liabilities).

There are a number of financial variables that are common across many economic surveys. By harmonizing the definitions of these variables and systematically applying standards, common content has been developed and implemented across programs.

The IBSP content model is based on a series of generic modules that cover common variables, and are applied to surveys without modifications from one survey to the next. This approach plays a critical role in creating coherence across programs and in minimizing the amount of effort required to build, test and implement survey content.

The standardised modules are in essence a series of business survey questions used to collect information to meet stakeholder requirements. There are standardized modules for income statement data (revenues and expenses), sales data by type of client, sales data by client location, and purchased service inputs.

The objective of utilizing tax information to its full potential guided the development of questionnaire content. Specifically, IBSP revenue and expense variables have been mapped directly with information available on tax files. This direct link eliminates the need for collecting financial information from small and medium enterprises, since data for these can easily be accessed from administrative sources.

One key issue that had to be resolved in developing financial data content was ensuring that the conceptual needs of the Canadian System of National Accounts are met through the use of administrative data. The COA bridges the two sets of concepts. As part of developing the IBSP content model, the COA was reviewed and revised to ensure that COA variables, which are directly linked to tax concepts, meet the information requirements of national accountants.

To add flexibility and meet specific survey requirements, subject matter staff can customize certain modules appearing on their IBSP questionnaires. For example, products appearing on manufacturing questionnaires will be different from those appearing on service industry questionnaires. And some of the standardized modules might not be required because they are not relevant for the industry. In constructing IBSP survey questionnaires, staff simply select relevant standardised content modules and then focus efforts on developing industry specific content where required. This greatly reduces the time needed to develop, implement and test new questionnaires.

3. Refer to Martineau (2012) paper in references.

6. Managing Response Burden

There is a rich history of using administrative data at Statistics Canada to reduce response burden. In fact, under the UES business survey model, tax data were used as a direct substitute for a sub-sample of sampled units and for imputation of non-response records.

Over time, through its use, tax data imputation methods improved and the quality of information increased which led to an even greater reliance on tax data as a primary information source. In the transition to the IBSP model, methods were adapted to take full advantage of tax data availability, which will result in additional response burden reductions across survey programs.

a) Tax replacement strategy

Tax data are the backbone of financial data estimates for IBSP surveys and the majority of sampled businesses will not be required to provide data for revenue and expense income statement modules.[4] By relying on administrative data as the primary source of financial information, the IBSP questionnaires will be more geared to measure information such as commodities produced, business practices and other characteristics including research and development activities and types of capital expenditure.

The use of administrative data is fully integrated in the IBSP sampling design. For each IBSP survey, the population of interest is divided in two parts:

- the very complex enterprises (there are approximately 2,000 on the BR) for which tax data cannot easily be used and;
- small and medium enterprises where tax data are used directly for the vast majority of financial variables.

For the most part, large and complex enterprises will be surveyed for all the information required, including financial and characteristic data. This is a requirement since Statistics Canada is mandated to publish financial information by industry and geographic region. Administrative information is often not available at this level of detail for complex enterprises that have significant activities in more than one industry and/or more than one province.

However, to reduce response burden for some of these very complex enterprises, a tool was developed to allow them to report only once through a combined questionnaire for all of their activities. The large and complex reporting tool is described in section 7. In addition, some complex enterprises will be tax replaced through the IBSP smart replacement strategy described below.

For simple enterprises, sample rotation and coordination methods will be introduced to ensure that response burden is well managed.

The sample design is complemented by estimators and associated quality indicators that take into account the design to produce the best estimates possible with the available data.

b) Smart replacement strategy

In the past, all large and complex businesses were exempt from tax replacement strategies. Under the IBSP smart replacement strategy, it was determined that administrative data can be used to replace survey data for some complex businesses if the administrative data satisfy criteria including availability, accuracy and coherence[5] and if the enterprise displays a certain degree of operational and structural stability for the most recent 3 years.

This is a major paradigm shift. IBSP surveys will only collect financial information from large and complex businesses if the above mentioned conditions are not met. To implement this strategy, each complex business must be reviewed independently to determine if it is a candidate for smart replacement.

4. Collection of survey income statement data will still be required for certain large and complex business and in cases where micro data are shared with other federal departments due to tax legislation statutes.
5. Refer to Martineau (2013) paper in references.

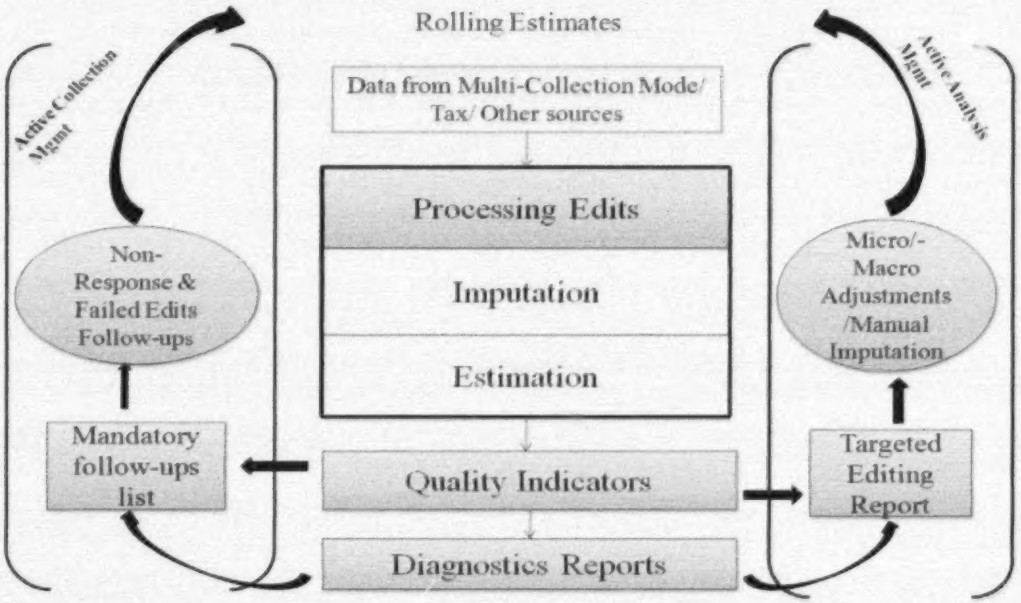
Administrative data are evaluated through comparisons to survey data for the three most recent survey cycles. This review includes a data accuracy comparison, a review of the stability of the administrative data source, and a review of the breadth of data available from the administrative source. The coherence team in Enterprise Statistics Division is responsible for preparing cases for review, presenting smart replacement candidates to subject matter specialists, and implementing the strategy for selected businesses.

c) Active collection management

Active collection management is a term used to indicate that collection efforts are dynamically adjusted based on data already received.

The cornerstone of active collection management is the Rolling Estimates (RE) model.[6] In this model, estimates are produced in an iterative fashion until an acceptable level of quality is reached. Quality indicators for key variables by domain of estimation and measures of impact for each unit will provide the information necessary to determine which outstanding units need to be followed up and which do not. In addition to collection management, the rolling estimates also play a key role in the IBSP analytic strategy which is described later in this paper.

Figure 2
The Active Collection Management process



If all quality targets are met for a specific survey, active collection will be closed and follow-up can be stopped; otherwise, follow-up or editing resources will be allocated to units that are deemed influential to key estimates and their quality. The remaining units will be run through automated processes to produce a consistent and usable micro-data file.

Expected benefits of the Rolling Estimates model and active collection management are an overall reduction in collection follow-up activities, thus reducing response burden, and a more coherent strategy for targeting quality across the different domains of estimation.

6. Refer to Mills et al. (2013) paper in references.

7. Incorporating a Coherence Analysis Framework

The vast majority of IBSP outputs will be integrated by the System of National Accounts into their various frameworks (e.g., Input-Output tables) and compared with each other by other stakeholders. Harmonizing concepts across business surveys certainly helps to make data from various programs more coherent, but this in itself is not sufficient. It is also extremely important to conduct coherence analysis early in the survey processing cycle to identify data issues, including gaps between enterprise and establishment level data, duplication across different survey programs, and instances where concepts are not reported consistently.

In the UES model, coherence analysis generally took place at the end of the survey cycle and often after data had already been disseminated. The process involved comparing reported data for selected businesses across multiple surveys and also to administrative data sources. The preliminary analysis was conducted by Statistics Canada's coherence team and results were shared with staff in subject matter divisions and staff in the System of National Accounts (SNA).

This work proved extremely valuable for identifying statistical and reporting issues and determining solutions. However, because of the timing of the process, it was difficult for subject matter divisions to incorporate changes into their statistical output. Often, the data had already been released and subject matter areas were forced to wait until planned data revision dates before incorporating the coherence analysis results.

In the IBSP model, data will be available in processing systems almost instantaneously allowing for coherence checks at the beginning of the survey process, i.e., during active collection. By identifying issues early on, there is ample opportunity to confirm information with respondents and enhance the accuracy of reported data, thus minimizing the risk of significant revisions.

A prototype coherence analysis viewer was developed to facilitate this work. The viewer includes data from auxiliary survey and administrative sources. Including all available information for a given business in one easily accessible placeholder facilitates a more thorough coherence analysis process, and will result in higher quality data outputs.

8. *Large and Complex Enterprises – A Special Case*

The majority of business and agricultural surveys conducted by Statistics Canada are based on samples, which means that only a subset of units in the target population are selected to complete a questionnaire. However, since samples are designed to attain high quality estimates while minimizing response burden, the largest units in terms of revenues generated are always selected. This is because the largest businesses and farms significantly impact statistical estimates. By obtaining data from these units, it is possible to achieve quality targets while sending comparatively few questionnaires to small and medium-sized businesses and farms.

For this strategy to be effective it is imperative that the largest firms consistently report their information, which is a challenge since they are asked to participate in numerous surveys. In an effort to facilitate sound reporting and to enhance relations with these firms, Statistics Canada has developed and implemented two programs.

a) EPM-LAOS programs

The first is the Enterprise Portfolio Manager Program (EPM), which is mandated to work with Canada's largest and most complex businesses to keep frame profiles of their operations up to date, to negotiate reporting arrangements, to ensure coherent data reporting across surveys, and to provide a single point of entry to rapidly respond to concerns or issues raised by the businesses in the program or by industry analysts.

The second program is the Large Agricultural Operations Statistics Program (LAOS). Like the EPM program, the LAOS team is responsible for ensuring that profiles for the largest farms in the country are up to date and reflect all current farming operations. In addition, the team is responsible for the collection of data from these farms including the development of customised data reporting arrangements.

Both programs have been highly successful. The EPM program covers 330 enterprise groups which together have several thousand operations as defined by establishments on the Business Register. Response rates consistently reach 90% for the businesses covered by this program.

The LAOS programs covers 370 enterprises which are comprised of about 800 agriculture operations that farm on 8,000 parcels of land. Response rates of 90% are generally attained.

b) Customized collection

During the past several years, both the EPM and LAOS teams have worked extensively to develop customised reporting arrangements for several of the units covered by their respective programs. These arrangements have been made to address respondent concerns that the statistical information requirement was overly burdensome and that there was duplication across various programs.

Since many large businesses and farms were being selected for monthly, quarterly and annual surveys, there certainly was an argument to be made for sharing information across programs. To achieve a reasonable compromise of what data should be collected by each survey program, and what could be shared across programs, the EPM/LAOS manager generally meets with subject matter analysts to elaborate a strategy.

Strategies developed often include the creation of a reporting spreadsheet that incorporates essential variables to be collected. The spreadsheet approach has helped to achieve the goal of improving relations with respondents by reducing reporting burden and eliminating duplication. However, such reporting strategies also have resulted in substantial manual effort to integrate reported data into processing systems. This resource intensive work is not integrated across programs and, as a result, coherence analysis cannot take place early in the survey cycle. It can only be done once all programs have incorporated the data into their systems.

Under IBSP these issues will be resolved by implementing a newly developed customisable reporting tool. The tool will be similar to the spreadsheet tool that is currently used for select complex reporting arrangements. However, there will be a number of significant improvements. The tool has been designed with metadata that will allow reported information to be seamlessly transferred to the collection infrastructure.

This will not only save time but will reduce the potential of data capture errors. Collection edits will be applied in a timely fashion allowing EPM and LAOS staff to follow-up with their respondents soon after the data arrives. And the coherence team will be able to identify and resolve issues much earlier in the survey cycle which will improve the quality of survey estimates.

9. Methodology Approach

The IBSP methodological approach takes into account core IBSP objectives including reducing response burden, maintaining quality, and maximizing the use of administrative data. Two of the key methods implemented to meet these objectives are the two phase sampling approach and the calibration of estimates to tax data sources.

a) Two-phase sampling

The two-phase sample design[7] is intended to improve quality while also minimizing response burden. IBSP surveys target variables such as the types of commodities produced or specific activities, such as research and development. In the UES model, sample units were selected based on NAICS and financial variables only, which made it difficult to control for the quality of commodity estimates. This was a shortcoming of the model since commodity estimates are crucial inputs for compiling National Accounts.

The two-phase sampling approach involves expanding the pre-contact step which is commonly in place for economic surveys. Specifically, the idea is to add questions to the collection application to obtain information that can be used later in sample selection. Examples include obtaining information about a business's capital expenditures, research and development activities, and the types of commodities produced.

7. See Turmelle et al. (2013) paper and Turmelle et al. (2012) paper in references.

The aim is to produce better estimates at a lower cost for variables or sub-populations that are not easily identifiable in the Business Register. The phase 1 collection application's flexibility will also make it possible to include other types of questions in current or cost-recovery surveys in accordance with users' needs, a key element of increasing flexibility in the IBSP model.

The decision whether to use the two-phase approach is made on a survey by survey basis. The annual surveys of manufacturing, capital expenditures, research and development, non-store retail trade, wholesale trade, energy and agriculture could benefit from this approach to respond rapidly to certain events. For example, only 40% to 60% of the 30,000 units selected in the historic Capital Expenditure Survey sample actually report capital expenditures. The two-phase approach will help reduce response burden and lower costs by identifying which businesses to target with the detailed capital expenditure questionnaire in phase 2.

After phase 1 is completed, the information collected will be processed in preparation for phase 2 sample selection. Throughout phase 1 collection, subject matter analysts will be able to identify updates via the Business Register. BSMD will incorporate the updated information to generate the phase 2 sample.

At this stage, survey collection begins, with electronic questionnaires serving as the main instrument. NAICS updates obtained from collection will be available directly from the BR while data collected from survey supplements, covering establishment locations, will be available via the IBSP warehouse. Once validated updated information from the supplements will be input to the BR by either the Enterprise Portfolio Management Program or the BR profile program.

b) Calibration

The estimation strategy for IBSP takes into account the two-phase sample design and incorporates calibration estimation to produce high quality program estimates.

In the two-phase sample design, there are two occurrences of sample selection. A unit can be selected during the phase 1 selection and then, if selected, it could be selected during the phase 2 selection. Thus there are two sampling weights, w_{1k} for phase 1 selection and w_{2k} for phase 2. However, during estimation the two weights are combined to produce a single weight, denoted as w_k, for each unit in phase 2.

In the IBSP, a calibration estimator is then used to calibrate the estimates with tax data totals. Calibration uses auxiliary data to produce an estimator that has a smaller variance than non-calibrated estimators and that will benchmark to the totals of the auxiliary data. This is achieved by using the auxiliary data to modify the sampling weights into what are known as calibration weights. The key point is that calibration produces a calibrated weight, $\tilde{w}_k$ that replaces the sampling weight, w_k, at the estimation step.

These calibrated weights are calculated only for take-some units. The take-all units will remain with a weight of one. For the take-none units, tax data, or values estimated through the use of tax data, will be aggregated and added to the take-all and take-some components to compile the full universe estimate.

10. Strategy for Analysis and Dissemination

The IBSP model was designed to facilitate a top-down approach for data analysis and validation. For many programs, this approach is a departure from methods previously employed. The idea of a top down approach is to produce and review estimates at a very early stage in the survey processing cycle and then to direct resources to undertake work that will result in significant quality improvements to the estimates. The IBSP top-down approach will be applied as data are collected through active collection management, which was described in section 6 c), and through the common analytic strategy.

a) Analytic strategy

The key to the new analytic approach is the production of estimates at an early stage of survey processing. In the past, for many annual programs the first set of estimates was only produced a few months in advance of data being disseminated. It was only at this time that analysts would have a macro view of their industry estimates and be in a position to compare data from alternative sources as part of the validation process. Prior to the estimates being produced, analysts often spent a lot of time and effort to manually correct micro data. Although this manual work was perceived to result in much higher quality estimates, studies indicated that some of the effort did not impact estimates.[8]

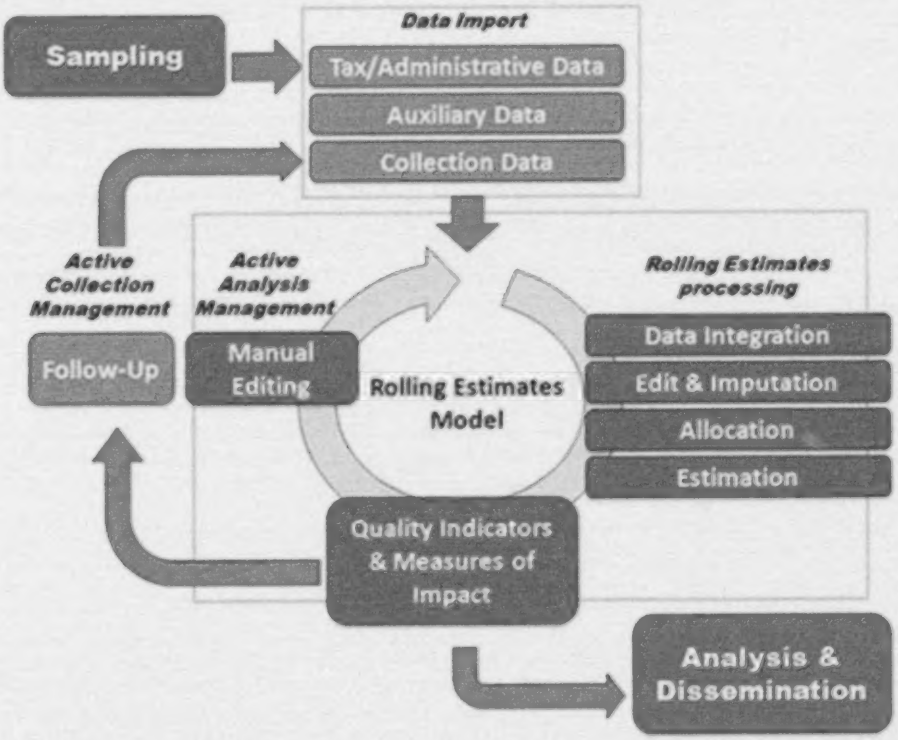
By producing estimates much earlier in the process, analysts will be able to more easily identify domains of estimation that are not coherent and can focus efforts on these domains. This focused editing approach is sure to have a quantifiable impact on the estimates. It may also result in timeliness gains since some domains will not require significant manual intervention to achieve quality targets.

To produce the rolling estimates, the IBSP system will access all available data for each sampled unit, including reported data, historical data, and administrative data. Non-response records will be imputed to complete the micro-data set. Allocation and estimation will then be run without manual intervention. Estimates will be available for review through the IBSP analytic tool.

Subject matter staff will compare aggregate estimates with auxiliary data and expected values, based on industry knowledge, to determine if and where further review and correction is required. A set of reports will be made available to identify high impact records in domains which require further review.

8. Refer to Saint-Pierre et al. (2011) paper and Godbout et al. (2011) paper in references.

Figure 3
The Rolling Estimates Process



b) Integration with the System of National Accounts (SNA)

Staff working in the SNA are key users of IBSP data. And many IBSP processes, methods and concepts were put in place to ensure that IBSP data can easily be integrated into the accounts.

For example, the adoption of harmonized content based on industry, commodity and financial standards results in more coherent information that is designed to mesh with the SNA conceptual framework. This will result in fewer transformations of data by SNA staff.

IBSP analytic and data access tools have been designed so that both subject matter and SNA staff can use the same interfaces to access and analyse data residing in the IBSP data mart. This will facilitate communications between the two groups since both have the same window into the data and access to all relevant information to help explain the data, such as comments from respondents or explanatory notes added by subject matter staff.

11. Governance Model

IBSP governance builds on the tried and tested structure that was successfully used to govern the UES over many years. That said, adjustments were made to accommodate the influx of new survey programs. With numerous new programs, several additional meeting participants had to be incorporated into the meeting framework.

The objective of the IBSP governance structure is to assure coherent and transparent decision making through a series of meetings that are action oriented. This is only achievable if staff attending meetings can represent their programs by making decisions. It is also necessary to identify the maximum number of participants within the various committees to ensure they can function efficiently.

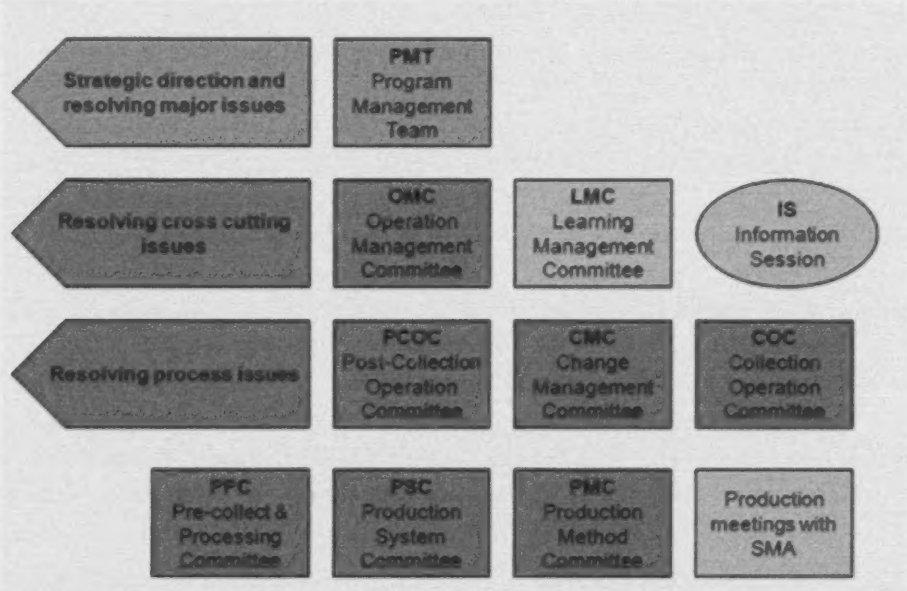
a) Committee structure

Figure 4 depicts the IBSP committee hierarchy. In total, there are eight interconnected committees that together govern all IBSP operational activities. Each committee has a specific mandate and decision making authority. At the top of the hierarchy, the Program Management Team is responsible for all strategic decisions. It is made up of senior executives who represent numerous subject matter and operational divisions.

Effective communications amongst the committees is paramount to keep the IBSP running smoothly. And the change management committee is a crucial link in the chain. It is through change management that key players are notified of issues and risks, and are asked to participate in issue resolution.

Clear communication is particularly important when changes to processes are required. The corporate JIRA tool has been implemented in the IBSP change management process and is the mechanism for keeping partners abreast of project issues and risks. But it is not just a tool to communicate information. The JIRA change and issue management system also provides staff with the ability to assign actions, document issues and solutions, and even escalate issues to senior staff if required.

Figure 4
Governance structure


Strategic direction and resolving major issues
PMT Program Management Team
Resolving cross cutting issues
OMC Operation Management Committee
LMC Learning Management Committee
IS Information Session
Resolving process issues
PCOC Post-Collection Operation Committee
CMC Change Management Committee
COC Collection Operation Committee
PPC Pre-collect & Processing Committee
PSC Production System Committee
PMC Production Method Committee
Production meetings with SMA

b) Management information system

Surveys integrating into IBSP will benefit from having access to a multitude of standardized reports to assess the status of processes and to determine if operational activities are on track.

There is an integrated schedule of activities with key milestones that is compiled and updated by staff in Enterprise Statistics Division, based on consultations with subject matter and operational partners. The schedule is an important component of the management information system. It is used to ensure activities are properly aligned and to identify potential bottlenecks when activities are not completed as forecast.

In addition to the schedule there are a series of information reports including:

- collection reports to display progress in terms of response rates and the status of edits;
- quality assessment reports to ensure data flowing between processes are complete and adhere to requisite levels of quality;
- edit and imputation reports that provide subject matter analysts with information about the quality of data after the edit and imputation process;
- a series of analytic reports that are used to compare data over time and between process iterations;
- reports to assess the sampling and estimation processes and the soundness of rolling estimates;
- quality indicator reports that are used to help analysts focus on records that have a high impact on estimates.

The IBSP information management system is also built with flexibility to add or modify reports based on program requirements. It will be accessible by all partner divisions to ensure everyone has the latest available information to make informed decisions.

12. Conclusion

The IBSP project has been challenging on many fronts, but none more so than the effort required to negotiate the application of common generic strategies to programs that traditionally utilized custom built solutions. A combination of strong governance, transparency in decision making and the involvement of partners in the development of the solutions was a successful approach. Frequent communication at all levels and through many different channels kept stakeholders informed and engaged in the project.

The IBSP has introduced many changes in all activities associated with collecting data, processing data, and compiling estimates. Everything from content review to the way the different areas involved in the work interact with each other has been modified. Without doubt, over the next few years, there will be a continuous need to refine as we learn more about what works well and what needs adjusting. As in the case of the UES program, it will take a number of years before the program will mature and reach a steady state. To get there, the continuous engagement of all stakeholders will be key, as well as the support of senior management.

The IBSP will provide many benefits. Updated survey methods and processes will take advantage of the latest available systems technologies. Data will be more coherent across programs through the application of harmonized concepts and standardized questionnaire content, making it easier to integrate outputs into the System of National Acconts. Response burden will be reduced through more intensive use of administrative data. Respondents will have an electronic reporting option and less collection follow-up will be required. Analysis will be facillitated through the introduction of rolling estimates and the application of common tools for validating data. And many additional programs will benefit from being incorporated into the IBSP model.

13. References

Brodeur M., Koumanakos, P., Leduc, J., Rancourt, E., Wilson, K. *(2006)*. The integrated approach to economic surveys in Canada, Statistics Canada. Catalogue 68-514.

Enterprise Statistics Division. (2010). Integrated Business Statistics Program Blueprint, Statistics Canada internal document.

Godbout, S., Beaucage, Y., Turmelle, C. (2011). Achieving Quality and Efficiency Using a Top-Down Approach in the Canadian Integrated Business Statistics Program, Paper for the Conference of European Statistics, Slovenia, 2011.

Hostetter, E. (2013). IBSP Conceptual Framework, Statistics Canada internal document.

Martineau, P. (2013). Smart Tax Replacement, Statistics Canada internal document.

Martineau, P. (2012). Use of the Chart of Accounts in determining the content of Statistics Canada business surveys, Statistics Canada internal document.

Mills, F., Godbout, S., Bosa, K., Turmelle, C. (2013). Mulivariate Selective Editing in the Integrated Business Statistics Program, Joint Statistical Meeting.

Saint-Pierre, É., Bricault, M. (2011). The Common Editing Strategy and Data Processing of Business Statistics Surveys. Paper for the Conference of European Statistics, Slovenia.

Sambell, C., Maffett, P., Felx, P. (2012). The Processing of Rolling Estimates, Findings and Issues, Statistics Canada internal document.

Turmelle, C., Beaucage, Y. (2013). The Integrated Business Statistics Program, Using a Two-Phase Design to Produce Reliable Estimates, Statistics Canada's Symposium, 2013.

Turmelle, C., Godbout, S., Bosa, K. (2012). Methodological Challenges in the Development of Statistics Canada's New Integrated Business Statistics Program, International Conference of Establishment Surveys, 2012.